CENTRE
SOPHIA ANTIPOLIS

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
B.P. 105
78153 Le Chesnay Cedex
France
Tél.: 954 90 20

Rapports de Recherche

N° 69

SUR LES SYSTÈMES DYNAMIQUES LINÉAIRES IMPLICITES SINGULIERS

Pierre BERNHARD

Avril 1981

ON SINGULAR IMPLICIT LINEAR DYNAMICAL SYSTEMS

Pierre BERNHARD December 1980

Summary

We investigate properties of existence, unicity, representation, of the (causal) solutions of implicit linear systems (or "generalized systems") when the underlying matrix pencil is singular. We relate the geometric and the algebraic approaches. The main conclusion is that if the underlying matrix pencil is "column singular" (i.e. has a non empty set of column minimal indices) the causal solutions, when they exist, can exactly be represented as the output of a classical two player dynamical system, where the second player accounts for the non unicity. Properties of the equivalent system are related to those of the singular matrix pencils made with the given matrices.

SUR LES SYSTEMES DYNAMIQUES LINEAIRES IMPLICITES SINGULIERS

Pierre BERNHARD Décembre 1980

Résumé

On étudie les propriétés d'existence, d'unicité et de représentation des solutions (causales) de systèmes linéaires implicites (ou systèmes généralisés) quand le faisceau de matrices sous-jacent est singulier. Nous faisons le lien entre les approches géométrique et algébrique. La conclusion principale est que si le faisceau de matrices est "colonne-singulier" (a un ensemble non vide d'indices minimaux de colonnes), les solutions causales, quand elles existent, peuvent être exactement représentées comme la sortie d'un système linéaire classique à deux joueurs, où le deuxième joueur représente la non unicité. Les propriétés du système équivalent sont reliées à celles des faisceaux de matrice singuliers construits à partir des matrices données.

1. - INTRODUCTION

1.1. - Problems considered

We study systems given in one of the following two forms, respectively discrete and continuous :

(*) $$E\, y(t+1) = F\, y(t) + G\, u(t),$$

(**) $$E \frac{dy}{dt}(t) = F\, y(t) + G\, u(t),$$

with the following definitions :

$y(t) \in \mathbb{R}^m$ is the (fundamental) output of the system,

$u(t) \in \mathbb{R}^p$ is the input

E and F are $r \times m$ constant matrices, G is a $r \times p$ constant matrix. r is called the rank of the system. It may be larger than, equal to or lower than m.

The questions of existence and unicity we shall investigate arise only if E is not invertible (in case $r = m$). We shall also consider problems of representation and canonical forms. We are mainly interested in singular systems, where the solution is non unique.

DEFINITION 0 - If $r = m$, the system is called square.

PROPOSITION 0 - A system (E, F, G) is always equivalent

i) to a system with rank equal to the rank of the composite matrix [E F G].

ii) if this rank is lowerthan m, to a square system.

PROOF -i) If the lines of the composite matrix [E F G] are not independant, we can always delete redundant equations in (*) or (**).

ii) If $r < m$, we can add lines of zeros to them.

HYPOTHESIS 0 - Because of property i) above, we shall always assume that rank $[E\ F\ G] = r$.

1.2. - Motivations

i) P.I.D. control. Systems of the form (**) naturally arise when applying output derivative feedback to an ordinary system. There the resulting implicit system is square. The interesting question is its limit behaviour when the E matrix is "close" to be singular. A perequisite to a complete understanding of the resulting "infinite frequency" modes (see [1]) is the present analysis.

ii) Systems with a linear state or state-control constraint. An equation of the form

$$0 = C\,y + D\,u$$

may be added to a standard system as an extra set of equations, resulting in a matrix E made of the identity and lines of zeros. There $r > m$.

iii) Interconnected systems. The natural statement of the equations of sets of interconnected systems may lead to equations of the type (ii).

iv) Econometric systems. Econometric systems are almost always of the form (**) (or a more complex one with non linear r.h.s.). Most famous among them are Leontief's models, and ARMA models with non invertible leading coefficient.

v) Perturbed systems. The perturbed system

$$\dot{x} = A\,x + B\,u + C\,v$$

is equivalent to the implicit system

$$E\,\dot{x} = E\,A\,x + E\,B\,u$$

where E is a matrix of maximum rank such that $E\,C = 0$.

vi) Time reversibility in discrete time systems. Backward projection for a standard discrete systems

$$x_{k+1} = F\,x_k + G\,u_k$$

leads to the study of the backward system

$$F\,\bar{x}_{k+1} = \bar{x}_k - G\,\bar{u}_k$$

where $\bar{x}_{k+1} = x_{-(k+1)}$ and $\bar{u}_k = u_{-k-1}$.

vii) Operator splitting numerical methods. Solution of the equation

$$A\,y = f$$

can be pursued using a recursion of form (*) with $A = E - F$ and $G\,u = f$ = constant. (or $G\,u_k \to f$).

viii) Implicit differential equations. The representation results obtained here may be of some interest in their own sake in the study of implicit linear differential equations.

1.3. - Originality

More than ten years ago, Rosenbrock's theory was explicitly devised to address implicit systems, of a more complicated type since higher derivatives were allowed as well as derivatives of the control. See a rather complete account in ROSENBROCK [2]. Since then, the precise type of systems we study have been investigated by Luenberger and co-workers ([3], [4], [5]). Beyond problems of existence and unicity, they have considered optimization problems. More recently papers by Verghese, Kailath and coworkers have dealt with the infinite frequency aspects of these systems [1], [6]. Systems of the form (*) also appear in connection with linear programming, see for instance [13].

All the above references deal with the "regular case", i.e. square systems with $\det(z\,E - F) \not\equiv 0$. In that case, as we shall see, existence implies unicity. Our main emphasis is on the singular case, and the representation of non unicity. Some works on that topics are due to Campbell. If [7] again deals only with the regular case, [8] considers a very particular instance of the singular case. It is a sub-case of our "static non unicity". Moreover, his application to linear systems is further restricted to the regular case.(1)

1.4. - Outline

In section 2, we develop the (elementary) geometric theory of strictly causal discrete systems (*). In the very short section 3, we check that all the results, but a minor one, carry over to the continuous case. In section 4 we investigate the geometric theory of the causal (but not strictly causal) case. Section 5 is devoted to the algebraic theory, invariants, transfer functions and canonical forms.

Footnote p. 4

(1) While this paper was being typed, the author got aware (through D. GABAY, of INRIA) of the work of WILKINSON [14]. It deals with the general singular case, but lacks the necessary tools of control theory to give a complete description of the nonunicity via invariants. It essentially covers the method of our paragraph 5.4, without the references to the geometric and transfer function theories.

2. - DISCRETE TIME SYSTEMS, THE STRICTLY CAUSAL CASE

2.0. - Causality

We quickly review here what causality, or strict causality, means for a dynamical system with possibly non unique solutions. We deal with the discrete system (*), the extension to (**) is straightforward, provided, in the definition of causality, "$\forall t$" be replaced by "for allmost all t". As a consequence, the difference between causality and strict causality, as given here, vanishes. Strict causality, in the continuous case, will carry an added requirement. See section 3.

Let Ω be the set of admissible control functions, i.e. applications from $[t_0,t_1]$ into $\mathbb{R}^m$. (Usually, $t_1 = +\infty$). A correspondence of solutions is a set-valued function S from Ω into the set of trajectories, which to each $u(\cdot)$ in Ω associates a set $S(u(\cdot))$ of trajectories $y(\cdot)$ satisfying (*). Let $S_\tau(u(\cdot))$ be the set of the restrictions to $[t_0,\tau]$ of the elements of $S(u(\cdot))$. We recall the

DEFINITION : The correspondence S is called strictly causal if, given $u_1(\cdot)$ and $u_2(\cdot)$ in Ω,

if $u_1(t) = u_2(t)$ $\forall t < \tau$, then $S_\tau(u_1(\cdot)) = S_\tau(u_2(\cdot))$.

S is said causal if the conclusion holds provided $u_1(t) = u_2(t)$, $\forall t \leq \tau$. (In all the sequel, "strictly causal" may correspondingly be replaced by "causal"). The set of strictly causal solutions of the system is the maximal strictly causal correspondence of solutions, i.e. the union $\bar{S}$ of all of them.
Given $u(\cdot)$ in Ω, a trajectory $y(\cdot)$ is called a strictly causal solution if it belongs to $\bar{S}(u(.))$.

A characteristic property of a strictly causal solution is that, in addition to satisfying (*) for all t, it is such that, for all τ in (t_0,t_1), the system (*) initialized at $y(\tau)$ has strictly causal solutions for every sequence $\{u(t),\ t \geq \tau.\}$

(The reader may easily check that this inductive characterization is indeed necessary and sufficient).

2.1. - Existence

We write $\mathcal{E} = \mathcal{R}(E)$ and $\mathcal{G} = \mathcal{R}(G)$ the respective ranges of E and G, as subspaces of $Y = \mathbb{R}^r$. Consider the following relation for a linear subspace $\mathcal{V}$ of Y :

$$(1) \qquad F\,\mathcal{V} \subset E\,\mathcal{V}$$

DEFINITION 1 - We call caracteristic subspace of the pair (E, F) the largest subspace $\mathcal{V}^*$ satisfying (1).

PROPOSITION 1 - This subspace exists, since {0} satisfies (1), and this equation being stable under addition of subspaces, $\mathcal{V}^*$ is the sum of all subspaces that satisfy it. (However, $\mathcal{V}^*$ may be trivial).

THEOREM 1 - The system (*) has a strictly causal solution over an interval of arbitrary length, for any control sequence u(.), iff

$$(2) \qquad \mathcal{G} \subset E\,\mathcal{V}^*$$

$$(3) \qquad y(0) \in \mathcal{V}^*$$

PROOF -i) Necessity. Let t be given. In order for y(t+1) to exist, it is necessary that

$$F\,y(t) + G\,u(t) \in \mathcal{E} \quad ,$$

and since this must be true for all $u(t) \in \mathbb{R}^p$, this implies

$$\mathcal{G} \subset \mathcal{E} \quad \text{and} \quad y(t) \in \mathcal{V}^o = F^{-1}(\mathcal{E}).$$

In order for the last relation to hold for every u(t-1), we need

$$\mathcal{G} \subset E\,\mathcal{V}^o, \qquad y(t-1) \in \mathcal{V}^1 = F^{-1}(E\,\mathcal{V}^o).$$

Continuing this process, we construct the sequence $\mathcal{V}^k$ by

$$(4) \qquad \mathcal{V}^{k+1} = F^{-1}(E\,\mathcal{V}^k)$$

and we must have for all k,

$$G \subset E\, \mathcal{V}^k, \qquad y(t-k) \in \mathcal{V}^k.$$

Necessity follows from the following fact.

PROPOSITION 2 - The sequence $\mathcal{V}^k$ is decreasing and converges to $\mathcal{V}^*$ in no more than m steps.

PROOF - From E $\mathcal{V}^o \subset \mathcal{E}$ follows that F $\mathcal{V}^1 \subset \mathcal{E}$, and thus $\mathcal{V}^1 \subset \mathcal{V}^o$, and so on by induction. However subspaces can decrease only by loosing one dimension, which cannot occur more than m times in $\mathbb{R}^m$. Let k be the first index such that $\mathcal{V}^{k+1} = \mathcal{V}^k$. The sequence $\mathcal{V}^k$ becomes stationary from this point on, and (4) shows that $\mathcal{V}^k$ satisfies (1). Therefore $\mathcal{V}^k \subset \mathcal{V}^*$. This establishes the necessity of (2) (3), but not the proposition, which states that $\mathcal{V}^k = \mathcal{V}^*$. This can easily be proved directly, but follows also from the sufficiency of (2) (3) that we now establish.

ii) Sufficiency. Let V be a rectangular injective (full column rank) matrix such that $\mathcal{R}(V) = \mathcal{V}^*$. (let dim $\mathcal{V}^* = n^*$, $V : m \times n^*$). Relations (1) and (2) imply

(5) $$\exists \bar{A} : F\, V = E\, V\, \bar{A}$$

(6) $$\exists \bar{B} : \quad G = E\, V\, \bar{B}$$

where $\bar{A}$ is a $n^* \times n^*$ matrix, and $\bar{B}$ is $n^* \times p$. We also have that $y(t) \in \mathcal{V}^*$ is equivalent to

(7) $$\exists\, \xi(t) \in \mathbb{R}^{n^*} : y(t) = V\, \xi(t).$$

Now, (*) is equivalent to

(8) $$E\, V\, \xi(t+1) = E\, V(\bar{A}\, \xi(t) + \bar{B}\, u(t))$$

which, together with (3) has the obvious solution

(9) $$\xi(t+1) = \bar{A}\, \xi(t) + \bar{B}\, u(t)\,,$$

$$y(0) = V\, \xi(0)\,.$$

∎

REMARK 1 - When (2) is not satisfied, we may restrict u to belong to $\mathcal{U}_{ad} = G^{-1}(E\,\mathcal{V}^*)$. In the sequel, condition (2) may always be understood to mean that this reduction has been performed.

2.2. - Unicity

DEFINITION 2 - We call characteristic kernel of the pair (E, F) the subspace $\mathcal{N}$ defined by

(10) $$\mathcal{N} = \text{Ker } E \cap \mathcal{V}^* .$$

Let $\dim \mathcal{N} = q$.

DEFINITION 3 - The pair (E, F) is said C-regular (or, more accurately, column regular), if $q = 0$:

(11) $$\mathcal{N} = \{0\}.$$

THEOREM 2 - Under conditions (2) and (3), the solution to equation (*) is unique, for any u(.), iff the system (the pair E, F) is C-regular. Otherwise the non unicity is described by the arbitrary choice of the sequence v(.) in equation (14) below.

PROOF -i) Unicity. Equation (8) implies (9) only modulo the kernel of EV, which reduces to {0} only under condition (10).

ii) Non unicity. If $\mathcal{N} \neq \{0\}$, let us chose a decomposition of $\mathcal{V}^*$ of the form

(12) $$\mathcal{V}^* = \mathcal{M} \oplus \mathcal{N} .$$

To this decomposition, we may associate a partition of V of the form

(13) $$V = [M\ N], \qquad E\,V = [E\,M \quad 0].$$

Let us partition accordingly ξ, $\bar{A}$ and $\bar{B}$ in the following way :

$$\xi = \begin{pmatrix} x \\ v \end{pmatrix}, \qquad \bar{A} = \begin{pmatrix} A & C \\ \tilde{A} & \tilde{C} \end{pmatrix}, \qquad \bar{B} = \begin{pmatrix} B \\ \tilde{B} \end{pmatrix} .$$

By definition EM is injective, so that (8) is equivalent to

(14) $$x(t+1) = A\,x(t) + B\,u(t) + C\,v(t),$$

(15) $$y(t) = M\,x(t) + N\,v(t).$$

■

The non unicity is therefore described as the effect of an extra input in a classical linear system. We may apply to it the tools of two player control systems. In that respect, it is worthwhile to notice that V being injective, knowledge of y is equivalent to the knowledge of both x and v. (This is important, for instance, in discrete capturability theory [10]).

REMARK 2 - The matric C may of course be of less than full column rank. If this is the case, by a proper choice of basis we can write

$$C = [C_1 \quad 0],$$

accordingly partitioning v in $v' = (v_1' \; v_2')$. Then v_1 must be considered as parametrizing a <u>dynamic non unicity</u>, since its effect propagates forward in time through the dynamics, while v_2 parametrizes a <u>static non unicity</u>, since it appears only in the output equation (15) (Recall that N is injective).

The triple (A, B, C) is clearly non unique. It may be altered through a change of basis within $\mathcal{V}^*$. This leads to the following fact.

PROPOSITION 3 - The pair (A, C) is uniquely defined up to a transformation of Kalman's feedback group (see KALMAN [11]).

PROOF - A change of basis within $\mathcal{V}^*$ can be described as

i) a change of basis within $\mathcal{N}$, i.e. on v.

ii) a change of choice of $\mathcal{M}$ within $\mathcal{V}^*$. Let $\tilde{M}$ generate an alternate $\tilde{\mathcal{M}}$:

$$y = M\,x + N\,v = \tilde{M}\,\tilde{x} + N\,\tilde{v}.$$

The difference $v - \tilde{v}$ depends linearly on y, and is null when $y \in \mathcal{N}$ i.e. when x = 0. Therefore it depends linearly on x alone :

$$\tilde{v} = P\,x + v.$$

Using the fact that M is injective, this gives

$$\tilde{x} = Q\, x .$$

where Q can be calculated as a function of M, $\tilde{M}$, N and P. Therefore this is equivalent to a state feedback super imposed on v and a change of basis on x.

iii) a change of basis on x alone (which can, of course, undo the previous one).

∎

We shall study further the invariants of (A, C). However an interesting geometric one at this point will be provided by the following definition.

DEFINITION 4 - We call <u>neutral subspace</u> of the pair (E, F) the smallest subspace $\mathcal{V}_*$ that satisfies (1) and contains $\mathcal{N}$.

PROPOSITION 4 - Such a subspace exists, and is the intersection of all subspaces that contain $\mathcal{N}$ and satisfy (1), $\mathcal{V}^*$ being one such subspace.

THEOREM 3 - Two solutions of (*) corresponding to the same initial point and same sequence u(.) are equal modulo $\mathcal{V}_*$. $\mathcal{V}_*$ is image by V of the reachable space of the pair (A C) in (14).

PROOF - By substraction, two solutions of (*) corresponding to the same initial point and the same sequence u(.), have their difference δy that satisfy

$$\delta y(t) = V\, \delta\xi(t) = M\, \delta x(t) + N\, \delta v(t), \qquad \delta\xi(0) = 0,$$

$$\delta x(t+1) = A\, \delta x(t) + C\, \delta v(t), \qquad \delta x(0) = 0.$$

Therefore, $\delta x(t)$ belongs to the reachable space of the pair (A, C). Conversely, any solution of this system remains strictly causal and satisfies

$$E\, \delta y(t+1) = F\, \delta y(t), \qquad \delta y(0) = 0, \tag{16}$$

and can therefore be added to a solution of (*) and still remain a solution.

The fact that $\mathcal{V}_*$ is exactly the (image of) reachable subspace of the pair (A, C) will be a corollary of theorem 5 below.

∎

2.3. - Minimality

DEFINITION 5 - We call maximum subspace of the triple (E, F, G) the largest subspace $\mathcal{W}^*$ satisfying

(17) $$F\,\mathcal{W}^* + \mathcal{G} = E\,\mathcal{W}^*.$$

PROPOSITION 5 - The subspace $\mathcal{W}^*$ exists, it is a subspace of $\mathcal{V}^*$, and is the limit, attained in no more than m steps, of the sequence $\mathcal{W}^k$ defined by

(18) $$\mathcal{W}^o = \mathcal{V}^*, \qquad \mathcal{W}^{k+1} = E^{-1}(F\,\mathcal{W}^k + \mathcal{G}) \cap \mathcal{V}^*.$$

PROOF - Notice that since $F\,\mathcal{V}^* + \mathcal{G} \subset E\,\mathcal{V}^*$, we have

(19) $$E\,\mathcal{W}^1 = F\,\mathcal{W}^o + \mathcal{G} \quad \text{and} \quad \mathcal{W}^1 \subset \mathcal{V}^*.$$

It follows easily that property (19) holds at every step of the algorithm, shifting the indices of $\mathcal{W}$ by an equal number, and also that the sequence is decreasing. It therefore has a limit which satisfies (17), of which it is easy to check that it is the largest solution of (17) (which is stable by addition of subspaces).

THEOREM 4 - $\mathcal{W}^*$ is the largest subspace traversed by the asymptotic regime of ($*$), i.e., $\forall k \geq n$, the application $(y(0), u(.)) \longmapsto y(k)$ is surjective over $\mathcal{W}^*$, which is exactly its range.

PROOF - By construction (3) implies $y(1) \in \mathcal{W}^1$, and by induction $y(t) \in \mathcal{W}^t$, with surjectivity. This, with the proposition, proves the theorem.

∎

If this results characterizes in some sense the reachable subspace of ($*$), it is not the most interesting one. As a matter of fact, classical system theory teaches us that the reachable subspace of interest is that which is reachable from the state zero. We therefore proceed with the following.

DEFINITION 6 - The minimal subspace of the triple (E, F, G) is the smallest subspace $\mathcal{W}_*$ satisfying (1) and (2) and containing $\mathcal{N}$.

PROPOSITION 6 - $\mathcal{W}_*$ exists, and is the limit (in m steps or less) of the same recurrence as in (18), but initialized with $\mathcal{W}_o = \{0\}$.

PROOF - That $\mathcal{W}_*$ exists is a consequence of the fact that its defining properties are stable under intersection, and that $\mathcal{V}^*$(as well as $\mathcal{W}^*$) satisfy them. Now, consider the recurrence (18) intralized with $\mathcal{W}_o = \{0\}$.

$$\mathcal{W}_1 = E^{-1}(\mathcal{G}) \cap \mathcal{V}^* .$$

Because of (2),

$$E\,\mathcal{W}_1 = \mathcal{G} .$$

The sequence $\mathcal{W}_k$ is clearly increasing, and by induction satisfies the same sequence of equalities of the form (19) as $\mathcal{W}^k$. By construction, $\mathcal{N} \subset \mathcal{W}_k$ for all k. Therefore, it has a limit that satisfies (1), (2) and contains $\mathcal{N}$. That it be the smallest such subspace is the consequence of theorem 5 and lemma 1.

THEOREM 5 - $\mathcal{W}_*$ is the image by V of the reachable space of the system (14) where both u and v are taken as controls.

PROOF - By construction, the reachable space for y(t) from y(0) = 0 is the limit of the recurrence of proposition 5. Therefore it satisfies (1), (2), and contains $\mathcal{N}$. We end the identification of this limit with $\mathcal{W}_*$ with the following lemma :

LEMMA 1 - Any subspace $\mathcal{W}$ satisfying (1), (2) and containing $\mathcal{N}$ contains the reachable space of (*) from zero.

PROOF - $\mathcal{W}$ is a subspace of $\mathcal{V}^*$, since it satisfies (1). Let us assume that the matrix V has been chosen in such a way that a submatrix W generates $\mathcal{W}$:

$$V = [L\ W] .$$

Since $\mathcal{N} \subset \mathcal{W}$, we may chose W such that N be a submatrix of it.

We may therefore partition V further in $W = (\tilde{M}, N)$ and therefore

$$V = [L\ \tilde{M}\ N],$$

with $M = [L\ \tilde{M}]$. Now, (5) and (6) give

$$\text{(20)} \qquad F\,M = E\,M\,A, \qquad F\,N = E\,M\,C, \qquad G = E\,M\,B,$$

which further partitioned according the above partition of M gives

$$(21)\qquad F\,\tilde{M} = E\,L\,A_{12} + E\,\tilde{M}\,A_{22}, \qquad G = E\,L\,B_1 + E\,\tilde{M}\,B_2.$$

Now, by hypothesis $\mathcal{W}$ satisfies (1) and (2), so that there exists $\tilde{A}_1$, $\tilde{A}_2$ and $\tilde{B}$ such that

$$(22)\qquad F\,\tilde{M} = E\,\tilde{M}\,\tilde{A}_1, \qquad F\,N = E\,\tilde{M}\,\tilde{A}_2, \qquad G = E\,\tilde{M}\,\tilde{B}.$$

If we remember that $(E\,L \quad E\,\tilde{M}) = E\,M$ is injective, comparison of (21) and (22) yield

$$A_{12} = 0, \qquad B_1 = 0.$$

This is the standard form for a system whose reachable space is contained in $\mathcal{W}$.

∎

COROLLARY 1 - The neutral space $\mathcal{V}_*$ is the reachable space of the pair (A, C).

PROOF - Apply theorem 5 with $G = 0$.

∎

Let now W be a submatrix of V generating $\mathcal{W}^*$, and let

$$(23)\qquad W = [\hat{M}\;N].$$

As before, there exists $\hat{A}$, $\hat{B}$ and $\hat{C}$ such that

$$(24)\qquad F\,\hat{M} = E\,\hat{M}\,\hat{A}, \qquad F\,N = E\,\hat{M}\,\hat{C}, \qquad G = E\,\hat{M}\,\hat{B}.$$

If the system (*) is initialized at $y(0) \in \mathcal{W}_*$, we can always represent its solution as

$$(25)\qquad \hat{x}(t+1) = \hat{A}\,\hat{x}(t) + \hat{B}\,u(t) + \hat{C}\,v(t),$$

$$(26)\qquad y(t) = \hat{M}\,\hat{x}(t) + N\,v(t),$$

and this constitutes a minimal representation of system (14) (15) (possibly with a feedback on v if we have changed of choice for $\mathcal{M}$). It can therefore be considered as a <u>minimal representation of (*)</u>. It is unique up to a change of basis and a feedback on v(.).

3. - CONTINUOUS TIME SYSTEM

This short section is aimed at checking that all previous results, except theorem 4, which is not important in the theory, carry over to the continuous case. We keep same notations and same numbers to the theorems.

Strict causality is taken to mean causality plus the fact that to a measurable input corresponds an absolutely continuous output.

3.1. - Existence

THEOREM 1 - PROOF

i) Necessity. Let $\mathcal{V}$ be the subspace generated by those y's that can be reached by the system. Necessarily, $\dot{y} \in \mathcal{V}$, therefore $\mathcal{V}$ must satisfy (1) and thus be included in $\mathcal{V}^*$, and (2).

ii) Sufficiency. Perform exactly as in § 2.1, to end up with

$$\dot{\xi}(t) = \bar{A}\,\xi(t) + \bar{B}\,u(t).$$

3.2. - Unicity

THEOREM 2 - PROOF unchanged, except for the substitution of an arbitrary measurable time function v(.) to the arbitrary sequence.

THEOREM 3 - PROOF. The proof that two solutions corresponding to the same initial condition and the same control function u(.) differ at each time instant of an element of the image by V of the reachable space of the pair (A, C) is unchanged. The rest of the theorem relies on the next paragraph.

3.3. - Minimality

Theorem 4 does not carry over in a simple way. One can prove that

$$y(t) \in y_o + t\,y_1 + \dots + \frac{t^{K-1}}{(K-1)!}\,y_{K-1} + \mathcal{W}^*,$$

where K is the smallest integer such that $\mathcal{W}^{K+1} = \mathcal{W}^K = \mathcal{W}^*$, and y_k is a sequence satisfying the homogeneous discrete system (16). The proof is a direct consequence of the remark that

$$\dot{y} \in \mathcal{W}^1 \Rightarrow y(t) \in y_o + \mathcal{W}^1 \Rightarrow \dot{x} \in E^{-1}(F(\mathcal{W}^1 + x_o) + \mathcal{G}) \cap \mathcal{V}^* = \mathcal{W}^2 + x_1,$$

and then iterating.

THEOREM 5 - PROOF. Defining $\mathcal{W}_*$ as previously, the algebraic constructions of section 2 remain the same. Moreover, classical system theory teaches us that given a pair (A, [B C]) the reachable space is the same for the discrete time system and the continuous time system. Therefore $\mathcal{W}_*$ is still the reachable space of the system. ∎

Notice that without the parallel between continuous time and discrete time systems, theorem 5 would be far less trivial in the continuous time case, since the identification of the limit of the recurrence $\mathcal{W}_k$ with the reachable space relies on a direct study of the system (*).

The corollary carries over unchanged.

4. - THE NON STRICTLY CAUSAL CASE

We investigate here existence, unicity and representation of the solution of (*) when y(t) is allowed to depend on past u(s) and on u(t). As for system (**), the same (algebraic) results hold if causality is defined via the existence of a proper transfer function (see section 5), since y(.) may now be non differentiable. The remarkable fact is the similarity of the conditions obtained here and in section 2.

4.1. - Existence

THEOREM 6 - There exists a causal solution to (*) over any time interval, for any sequence u(.), if and only if

$$\mathcal{G} \subset E\,\mathcal{V}^* + F\,\mathrm{Ker}\,E = E\,\mathcal{M} + F\,\mathrm{Ker}\,E, \tag{27}$$

$$y_o \in \mathcal{V}^* + \mathrm{Ker}\,E = \mathcal{M} + \mathrm{Ker}\,E. \tag{28}$$

PROOF - i) Necessity. Let us arbitrarily write

$$y(t) = z(t) + \varepsilon(t),$$

where $\varepsilon(t) \in \text{Ker } E$,

$z(t) \in \mathcal{Z}$,

and $\mathcal{Z}$ is a subspace that we shall chose later on. By an appropriate restriction, we can manage to have $\mathcal{Z} \cap \text{Ker } E = \{0\}$, so that the above decomposition of y is unique. Equation (*) yields

$$E\, z(t+1) = F\, z(t) + F\, \varepsilon(t) + G\, u(t), \tag{29}$$

so that, given y(t) and u(t), E z(t+1) is uniquely determined and also z once we restrict $\mathcal{Z}$ to have no intersection with Ker E.

By the same type of induction as in paragraph 2.1, we readly see that we must have

$$F\, \mathcal{Z} \subset E\, \mathcal{Z} + F \text{ Ker } E, \tag{30}$$

$$\mathcal{G} \subset E\, \mathcal{Z} + F \text{ Ker } E, \tag{31}$$

in order for (29) to have a solution (z(t+1), ε(t)) once z(t), which depends upon the past, and u(t) are given. The result then follows from the following fact.

LEMMA 2 - The largest subspace satisfying (30) is $\mathcal{V}^* + \text{Ker } E$.

PROOF - Notice first that $\mathcal{V}^* + \text{Ker } E$ satisfies (30). Now let $\mathcal{Z}$ satisfy (30), and contain Ker E (since the maximal one does).
Write

$$\mathcal{Z} = \mathcal{V} + \text{Ker } E,$$

and

$$F(\mathcal{V} + \text{Ker } E) \subset E\, \mathcal{V} + F \text{ Ker } E.$$

This implies that

$$\forall\, a \in \mathcal{V}\,,\ \exists\, \tilde{a} \in \mathcal{V} \text{ and } b \in \text{Ker } E, \text{ such that}$$

$$F\, a = E\, \tilde{a} + F\, b.$$

Clearly, $\tilde{a}$ and b can be chosen depending linearly on a. Let therefore K generate Ker E. There exists a matrix of appropriate type, such that, for every $a \in \mathcal{V}$

$$Fa = E\,\tilde{a} + F\,K\,P\,a,$$

thus

$$F(I - K\,P)a = E\,\tilde{a} = E(I - K\,P)\,\tilde{a}.$$

Let therefore

$$\bar{\mathcal{V}} = (I - K\,P)\,\mathcal{V}.$$

Clearly

$$\bar{\mathcal{V}} + \text{Ker } E = \mathcal{V} + \text{Ker } E = \mathcal{Z},$$

but also

$$F\,\bar{\mathcal{V}} \subset E\,\bar{\mathcal{V}},$$

so that

$$\bar{\mathcal{V}} \subset \mathcal{V}^*, \qquad \mathcal{Z} \subset \mathcal{V}^* + \text{Ker } E.$$

This proves the lemma. Notice that to get the unicity of z(t), we must chose $\mathcal{Z} = \mathcal{M}$, a complement of Ker E in $\mathcal{V}^*$.

ii) Sufficiency. Let $\bar{K}$ be a matrix whose columns span Ker E, and M be as in section 2. Let

(32) $$y(t) = M\,x(t) + \bar{K}\,w(t).$$

Condition (27) implies that there exists matrices $\bar{B}$ and $\bar{P}$ such that

$$G = E\,M\,\bar{B} + F\,\bar{K}\,\bar{P}$$

Now, equation ($\star$) can be written equivalently

(33) $$EM\,x(t+1) = EM\,Ax(t) + FK\,w(t) + EM\,\bar{B}u(t) + FK\,\bar{P}u(t)$$

so that one possible solution of ($\star$) is, using again (32),

(34) $$x(t+1) = Ax(t) + \bar{B}u(t),$$

(35) $$y(t) = Mx(t) - \bar{K}\,\bar{P}u(t).$$

(Notice that A is defined using only E and F, as in section 2. However since the requirement on G has been changed one should not look for a relation between the matrices G, $\bar{B}$, $\bar{P}$ of this section and G, $\bar{B}$ in the previous ones).

(34) and (35) together provide a causal solution and end the proof. ∎

4.2. - Unicity

THEOREM 7 - The causal solution of (*) under conditions (27) (28) is unique, for each sequence u(.), iff the pair (E, F) is column regular.

PROOF - We want to find under what conditions equation (33) has a unique solution x(t+1), w(t), once x(t) and u(t) are given. As a matter of fact, if this is true, since x(0) is uniquely determined by y(0), x(1) and w(0) will be unique, and all succeeding y's by induction.

By taking the difference $\delta x(t+1)$, $\delta w(t)$ between two solutions, we are led to the investigation of the non zero solutions of

$$EM\ \delta x(t+1) = FK\ \delta w(t)$$

The only solution is zero if and only if

(36) $$\text{Ker } F \cap \text{Ker } E = \{0\},$$

and

(37) $$E\,\mathcal{M} \cap F\ \text{Ker } E = \{0\}.$$

This is so because EM is injective. Therefore for a non zero solution, either both sides are zero, but then (36) does not hold, or there is a non zero element in $E\,\mathcal{M} \cap F$ Ker E.

Notice that

$$\mathcal{V}^* = F^{-1}(E\,\mathcal{M}),$$

so that

(38) $$\mathcal{N} = F^{-1}(E\,\mathcal{M}) \cap \text{Ker } E.$$

Now, it can easily be checked that for two subspaces $\mathcal{A}$ and $\mathcal{B}$, and an arbitrary linear operator F, one has

$$F\,\mathcal{A} \cap F\,\mathcal{B} = F[(\mathcal{A} + \text{Ker } F) \cap \mathcal{B}]$$

Apply this to (38), noticing that $F^{-1}(E\,\mathcal{M}) \subset$ Ker F, it comes

(39) $$F\,\mathcal{N} = E\,\mathcal{M} \cap F\ \text{Ker } E.$$

Notice also that Ker F $\subset V^*$, so that

(40) $\quad N \supset \text{Ker } E \cap \text{Ker } F.$

From (39) and (40) we conclude that if (37) or (36) is violated, N is non trivial, i.e. the system is not C-regular.

Conversely, if N is non trivial, and if moreover (36) holds then since $N \subset$ Ker E, (36) implies

$$N \cap \text{ Ker } F = \{0\},$$

and therefore F N has same dimension as N, and (39) shows that (37) is violated.

∎

REMARK 3 - We may again make a distinction between two types of non-unicity as in remark 2. In the case (37) holds,(but not (36)) the non unicity in y involves only w(t) and does not propagate in time. The sequence x(.) is unique. The non unicity may be called "static". The dynamic non unicity is induced by non zero elements in E $M \cap$ F Ker E.

The fact that the unicity condition be the same as in the strictly causal case will be more fully explained by the algebraic theory. It is not a trivial consequence of the fact that it is, in both case, a study of non zero solutions of (16), since y ranges over a larger subspace here.

4.3. - Representation

Let us be more precise in representation (32), puting

$$\bar{K} = [N\ K],$$

(and with w having now a different meaning)

$$y(t) = M\,x(t) + N\,v(t) + K\,w(t).$$

We also have (recalling that F $N \subset$ E M)

$$G = E\,M\,B + F\,K\,P,$$

so that (33) can now be written

$$EM\,x(t+1) = EM(Ax(t) + Bu(t) + Cv(t)) + FK(Pu(t) + w(t)).$$

But now,

$$\mathcal{R}(E\,M) \cap \mathcal{R}(F\,K) = \{0\},$$

since any part of Ker E whose image by F is in $E\,\mathcal{M}$ belongs to $\mathcal{V}^*$, i.e. to $\mathcal{N}$, and moreover, since clearly Ker $F \subset \mathcal{V}^*$, F K is, as well as E M, injective. Therefore the only solution is

$$w(t) = -\,P\,u(t),$$

or, defining

$$-\,K\,P = D,$$

$$y(t) = M\,x(t) + D\,u(t) + N\,v(t),$$

and

$$x(t+1) = A\,x(t) + B\,u(t) + C\,v(t).$$

These equations will be summarized further (eq. (48) to (52)). Notice that those for the strictly causal case are identical to these where we set D = 0. Notice also that the same analysis applies to a representation of system (**).

5. - ALGEBRAIC THEORY

5.1. - Generalized spectrum and regularity

DEFINITION 7 - We call generalized eigenvalue of the pair (E, F), and associated generalized eigenvector a complex number z, and a non zero complex vector ξ of $\mathbb{C}^m$, such that

$$(z\,E - F)\xi = 0. \tag{41}$$

LEMMA 3 - Both the real part and imaginary part of a generalized eigenvector of (E, F) belong to $\mathcal{V}^*$. Under condition (2) this is also true of the first component (in $\mathbb{R}^m$) of a generalized eigenvector of the pair ([E 0], [F G]).

PROOF - Let

(42) $$z = \sigma + i\omega \quad , \qquad \xi = \eta + i\,\zeta ,$$

be a generalized eigenvalue and eigenvector of (E, F). The (41) yields

$$F\,[\eta \quad \zeta] = E\,[\eta \quad \zeta] \begin{pmatrix} \sigma & \omega \\ -\omega & \sigma \end{pmatrix}$$

Calling X the subspace generated by $[\eta \quad \zeta]$, this reads

$$F\,X \subset E\,X$$

and according to proposition 1, this implies $X \subset V^*$, hence the first claim. Keeping the notations (42), let $\varphi \in \mathbf{C}^p$:

$$\varphi = \chi + i\,\psi$$

constitute with ξ a generalized eigenvector of ([E 0], [F G]) :

$$(z\,E - F)\xi - G\,\varphi = 0$$

Using (6), and separating again real and imaginary parts, we get

$$F[\eta \quad \zeta] = E[\eta\ \zeta] \begin{pmatrix} \sigma & \omega \\ -\omega & \sigma \end{pmatrix} - E\,V\,\bar{B}\,[\chi\ \psi],$$

which gives

$$F\,X \subset E\,X + E\,V^*$$

Adding $F\,V^*$ to the left and using (1), this gives according to proposition 1 $X + V^* \subset V^*$, and thus $X \subset V^*$

∎

THEOREM 8.- The generalized spectrum of the pair (E, F) is finite if and only if this pair is C-regular (definition 3). Otherwise, the generalized spectrum is the whole set $\mathbf{C}$, and the rank deffect of $z\,E - F$ is at least q for all z (where $q = \dim N$).

PROOF - Equations (5) and (6) yield

(43) $$E\,V(z\,I_{n^*} - \bar{A}) = (z\,E - F)\,V$$

Assume (E, F) is C-regular. Then E V is injective. Let ξ be a generalized eigenvector, we know according to lemma 3 that there exists a vector ν of $\mathbb{C}^{n^*}$ such that

$$\xi = V\,\nu$$

Placing this in (41) and using (43), it comes :

$$E\,V(z\,I - \bar{A})\nu = 0$$

and since E V is injective, z has to be an eigenvalue of $\bar{A}$, of which there are, at most, n^*.

To the contrary, assume now the (E, F) is not C-regular. Then (43) yields, partitioning V and $\bar{A}$ as in § 2.2 :

(45) $$(z\,E - F)\,[M \quad N] = E\,M[z\,I_n - A \quad -C].$$

The matrix $[z\,I - A \quad -C]$ has q less lines than columns. Therefore it has for all z's a kernel of dimension at least q. V being injective, it gives rise to a kernel of dimension at least q for (z E - F).

∎

This theorem is the justification for the definition 3. As a matter of fact, a <u>pencil of matrices</u> (z E - F) is said to be column singular if its columns are not independant as polynomials in $\mathbb{R}^n[z]$, a characterization that coincides with theorem 8. The degrees of the vectors of a polynomial minimal basis [12] of its kernel are called the Kronecker minimal column indices of the pencil. They are invariant under pencil similarity [9]. It also justifies the following definition :

DEFINITION 8 - We cal <u>essential eigenvalue</u> of the pair (E, F) a complex number z such that

$$\text{rank}\,(z\,E - F) < m - q$$

(It is a root of an invariant factor of the pencil (E, F)).

As a corollary of lemma 3 and theorem 8, we have :

COROLLARY 1 - q is the column rank deffect of the matrix pencil $(z E - F)$, equal to m minus the size of the largest non identically null determinant in this matrix.

- If $r < m$, $\mathcal{V}^*$ is never trivial, the system never C-regular, $q \geq m - r$
- If $r = m$, $\mathcal{V}^*$ is never trivial, the system is C-regular iff $\det (z E - F) \not\equiv 0$
- If $r > m$, $\mathcal{V}^*$ is non trivial iff the matrix $(z E - F)$ is reductible, i.e. all $m \times m$ determinants have a common root (for this value of z, the columns of $(z E - F)$ are not independant in $\mathbb{C}^m$). The system is C-regular iff one of the $m \times m$ determinants is not identically zero.

PROOF - According to lemma 3, if there exists a generalized eigenvector, $\mathcal{V}^*$ is non trivial. Conversely, if $\mathcal{V}^*$ is non trivial, (43) shows that (E, F) has generalized eigenvalues : those of $\bar{A}$ at least. Now, a generalized eigenvalue is clearly a complex number z such that the columns of $(z E - F)$ are not independant in $\mathbb{C}^m$, i.e. no $m \times m$ determinant is different from zero. And if the generalized spectrum of (E, F) is $\mathbb{C}$, all $m \times m$ determinants are null for all z's, i.e. identically zero.

∎

5.2. - Invariants

We first extend a theorem by KALMAN [11] to non completely controllable systems.

PROPOSITION 7 - Let (A, C) be a (non completely controllable) system. A complete set of invariants under the feedback group is given by :

i) the control invariants of the controllable part.

ii) the invariant factors of the uncontrollable part.

PROOF - With a change of basis, we can take the pair (A, C) into the form

$$(45) \qquad A = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix} \qquad C = \begin{pmatrix} C_1 \\ 0 \end{pmatrix}$$

where (A_{11}, C_1) is completely controllable (and constitutes the controllable part).

Partitioning a feedback K accordingly in $[K_1, K_2]$, we see that

$$A - CK = \begin{pmatrix} A_{11} - C_1 K_1 & A_{12} - C_1 K_2 \\ 0 & A_{22} \end{pmatrix}$$

All changes of basis on x that preserve the special form (45) can easily be seen to be of the form $\tilde{x} = P x$ with

$$P = \begin{pmatrix} P_{11} & P_{12} \\ 0 & P_{22} \end{pmatrix}, \qquad P^{-1} = \begin{pmatrix} P_{11}^{-1} & - P_{11}^{-1} P_{12} P_{22}^{-1} \\ 0 & P_2^{-1} \end{pmatrix}$$

so that

$$P A P^{-1} = \begin{pmatrix} P_{11} A_{11} P_{11}^{-1} & X \\ 0 & P_{22} A_{22} P_{22}^{-1} \end{pmatrix},$$

with

$$X = (- P_{11} A_{11} P_{11}^{-1} P_{12} + P_{12} A_{22} + P_{11} A_{12}) P_{22}^{-1}.$$

So that we can always perform the following sequence of transformations, within the feedback group.

1) Chose a feedback K_1 so that $A_{11} - C_1 K_1$ and A_{22} have no common eigenvalue (always possible because (A_{11}, C_1) is completely controllable).

2) Make a change of basis of the above type with $P_{11} = I$, chosing P_{12} such that $X = 0$. (This is a Lyapunov equation, which has a solution thanks to the first step).

3) Using a change of basis within the controllable subspace only (i.e. $P_{12} = 0$, $P_{22} = I$), and a new feedback $[K_1 \;\; 0]$ and a change of basis on the input space, bring (A_{11}, C_1) into Brunovskii's canonical form (as in KALMAN [11]).

4) Use a change of basis within the space x_2 (i.e. $P_{11} = I$, $P_{12} = 0$) to bring A_{22} in Jordan form.

The form thus obtained is invariant within the feedback group, and is entirely specified by the invariants we mentioned.

THEOREM 9 - Given a pair (E, F), the corresponding system (A, C) is entirely characterized by :

i) the control invariants of the controllable part of (A, C), which coincide with the Kronecker minimal column indices of the pencil (z E - F).

ii) the invariant factors of the uncontrollable part of A, which coincide with the finite invariant factors of the pencil (z E - F).

PROOF - Because of propositions 3 and 7, the elements quoted for the pair (A, C) are indeed a complete set of invariants. There only remains to relate them to the corresponding quantities of the pencil (z E - F).

i) from KALMAN [11], we know that the control invariants of the pair (A_{11}, C_1) are the minimal column indices of the pencil $[z\,I - A_{11} \quad - C_1]$. From (45), it follows that they are the same as the minimal column indices of the pencil $[z\,I - A \quad - C]$. As a matter of fact, let

$$\nu(z) = \begin{pmatrix} \nu_1(z) \\ \nu_2(z) \\ \mu\ (z) \end{pmatrix}$$

be a polynomial vector in $\mathrm{Ker}[z\,I - A \quad - C]$, this is equivalent to

$$(z\,I - A_{11})\nu_1(z) - A_{12}\ \nu_2(z) - C_1\ \mu(z) = 0,$$

$$(z\,I - A_{22})\nu_2(z) = 0.$$

However, $z\,I - A_{22}$ is a regular pencil, and therefore $\nu_2(z)$ is identically null (since it is a polynomial, null for all z that are not in the spectrum of A_{22}). Thus $[\nu_1'(z)\ \mu'(z)]'$ is in the kernel of $[z\,I - A_{11} \quad - C_1]$. According to lemma 3, all generalized eigenvectors, and therefore the basis vectors of Ker (z E - F), can be written as

$$\xi(z) = V\ \nu(z).$$

Therefore, using (44), we see that to each $\xi(z)$ in Ker(z E - F) corresponds a $\nu(z)$ in $\mathrm{Ker}[zI-A \quad -C]$ and conversely. Moreover, V being injective, $\xi(z)$ and $\nu(z)$ are of same degree.

ii) We now show that essential eigenvalues of (E, F) are eigenvalues of A_{22}, with the rank deffect of A_{22} equal to that of $(z\,E - F)$, minus q. Let λ be an essential eigenvalue of (E, F), with a corresponding kernel of dimension q + k. According to lemma 3 and (44), $[\lambda I - A \quad - C]$ has a kernel of dimension q + k in $\mathbb{R}^{n^*}$, with $n^* = n + q$. Therefore, only n - k of its lines are independant, and this is a fortiori true for $(\lambda I - A)$. Thus λ is an eigenvalue of A, with an associated eigen subspace of dimension at least k. Now this property is independant of the particular choice of basis within V^*, and thus, according to proposition 3, invariant under feedback. Therefore this eigenvalue and eigen subspace are associated to the uncontrollable part of A.

Conversely, considering the form (45) of (A, C), we have seen that polynomial vectors in Ker $[zI-A \quad -C]$ have a zero block in the uncontrollable part of the state space. Thus to an eigenvalue of A_{22}, with an eigensubspace of dimension k, correspond k generalized eigenvectors (that we shall chose with zero blocks in the first and third parts), independant of each other and of any vector in Ker $(z\,E - F)$. Therefore this complex number is an essential eigenvalue with a column rank deffect at least q + k.

At this stage, we know that essential eigenvalues of (E, F) are eigenvalues of A_{22}, and that the number of Jordan blocks associated to it coincide. There remains to prove that they are identical in dimension. The technique is the same, using Jordan chains, and only heavier. We shall not go into too much detail. To a Jordan block of $(\lambda E - F)$ corresponds a Jordan chain $\xi_1, \xi_2, \ldots \xi_p$ satisfying

$$\begin{aligned}
(\lambda\, E - F)\ \xi_1 &= 0, \\
(\lambda\, E - F)\ \xi_2 &= E\ \xi_1, \\
&\ \vdots \\
(\lambda\, E - F)\ \xi_p &= E\ \xi_{p-1}.
\end{aligned}$$

Here, p is the size of the Jordan block. There remains to check that all the ξ_i's are in V^*, and can be chosen independant of the vectors of Ker$(z\,E - F)$ at $z = \lambda$. Hence there are q + 1 independant solutions to each of the above equations, and consequently, using a linear combination with total weight one, we can find one with a zero component in N. Consequently, there corresponds to it a Jordan chain of $\lambda I - A$. Independance modulo Ker $(z\,E - F)|_{z=\lambda}$ in $\mathbb{R}^m$ corresponds to independance in $\mathbb{R}^n$. Therefore, elementary divisors of $(z\,E - F)$ are elementary divisors of $(z\,I - A)$, fixed under feedback, and thus, according to Rosenbrock's feedback theorem, elementary divisors of $z I - A_{22}$. The converse proof goes exactly as above.

∎

The particularization of the above results to the fact that the eigenvalues of A_{22} coincide with the essential eigenvalues of (E, F) lead to the following definition and corollary :

DEFINITION 9 - The system (*) or (**), satisfying (2) or (27), is called <u>stable</u> if for every bounded input function u(.), there exists a bounded (causal) output function y(.) from any initial condition.

COROLLARY 1 - The implicit system is stable if and only if the essential eigenvalues of the pair (E, F) are stable (i.e., of modulus less than one, or simple and of modulus unity in case (*), and of negative real part, or simple imaginary in case (**)).

PROOF - If the condition of the corollary is met, the equivalent system is stabalizable with v with a linear feedback (or equivalently, can be chosen stable). Therefore, there exists bounded solutions x(.) from any initial condition, with a choice of a bounded function v(.). (zero if the system is chosen stable). Therefore, y(.) as given by (15) or (49) remains bounded for these solutions.

To the contrary, if the condition is not met, there is a mode, uncontrollable with v, which is unstable. Therefore, except for a strict subspace of initial conditions, the solution x(.) will diverge for all choices of v(.). And since the matrix M is injective, and has a range $\mathcal{M}$ in direct sum with the range $\mathcal{N}$ of N, y(.) as given by (15), or (49) recalling that u(.) is assumed bounded, will diverge as well for all (causal) solutions.

∎

REMARK 5 - It is impossible to request, for singular systems, that all solutions be bounded, in view of theorem 3.

REMARK 6 - One may of course define in the same way asymptotically stable implicit systems.

Finally, one can clearly define the feedback group for systems (*) or (**) exactly in the same way as for an ordinary system. It clearly preserves existence of a strictly causal solution.

DEFINITION 10 - The implicit system is minimal if the minimal subspace $\mathcal{W}_*$ coincides with the characteristic subspace $\mathcal{V}^*$.
Then (15) (16) is completely controllable. We have :

THEOREM 10 - Under condition (2), if the implicit system is minimal, a complete set of invariants under the feedback group is provided by the Kronecker minimal indices of the matrix pencil $[z\,E - F \quad -G]$, and they coincide with the control invariants of the system (A, [B C]).

PROOF - The proof is in two steps. First check that the feedback group on the implicit system, combined with the non unicity pointed out in proposition 3, translates exactly in the classical feedback group for (A, [B,C]), and that conversely the latter generates the former. This in an easy consequence of the fact that V is injective. We leave it to the reader to check. Then, using the fact that (A, [B C]) is by hypothesis completely controllable and Kalman's theorem, we have that its control invariants are a complete set of invariants for the implicit system.

The second step is to identify the control invariants of (A,[B C]), i.e., according to KALMAN [11] the column indices of $[zI-A \quad -B \quad -C]$, with the column indices of $[zE-F \quad -G]$. This is done in the same fashion as in theorem 9, i), using the second claim of lemma 3 and

$$[(z\,E - F)[M \quad N] \quad -G] = E\,M\,[zI-A \quad -C \quad -B].$$

∎

5.3. - Transfer functions

THEOREM 11 - There exists a (strictly) causal solution to the system (*) or (**) iff there exists a (strictly) proper rational matrix K(z) such that

$$(46) \qquad (z\,E - F)\,K(z) = G.$$

Let also L(z) be a proper (not strictly) rational matrix, of maximum rank, such that

$$(47) \qquad (z\,E - F)\,L(z) = 0.$$

Then all solutions of the implicit system are given by

$$Y(z) = K(z)\ U(z) + L(z)\ V(z),$$

where Y(z) and U(z) are the z-transforms of y(.) and u(.) respectively, and V(z) an arbitrary power series of z^{-1}, of appropriate dimension.

PROOF - Notice first that there exist complex (column) vectors $\ell_i(z)$ satisfying (47) iff the pair (E, F) is not C-regular. It is easy to see (see [9]) that they can be chosen polynomial, or, dividing each such column by the highest power of z present in it (since (47) is homogeneous), rational proper. If these degrees are chosen as small as possible, they are the column minimal indices or Kronecker indices of the pencil.

i) Necessity. We know that, if a strictly causal solution exists, it is represented by (14), (15), or, in the non strictly causal case, by the following set (that coincides with the former if we set D = 0) :

(48) $$x(t+1) = Ax(t) + Bu(t) + Cv(t),$$

(49) $$y(t) = Mx(t) + Du(t) + Nv(t),$$

with the definitions of the matrices A, B, C, D, M and N as

(50) $$F[M\ \ N] = [EMA\ \ EMC],$$

(51) $$G + FD = EMB,$$

(52) $$ED = 0, \qquad EN = 0.$$

Hence the claim of the theorem for Y(z) with

(53) $$K(z) = D + M(z\ I - A)^{-1}\ B,$$

(54) $$L(z) = N + M(z\ I - A)^{-1}\ C.$$

We can calculate

(55) $$(z\ E - F)\ K(z) = (z\ E - F)\ M(z\ I - A)^{-1}\ B - F\ D.$$

Now, (44) still holds, with V = [M N]. Taking the first blocks in both sides, it comes

(56) $$(z\,E - F)\; M(z\,I - A)^{-1} = E\,M,$$

and therefore, (55) with (51) yield (46). Similarity, we have with the second block in (44) (or with 50),

$$(z\,E - F)\; N = EMC,$$

and this together with (56) yields (47).

ii) Sufficiency. Assume the two proper rational matrices K(z) and L(z) exist, satisfying (46) and (47). Consider the rational matrix

(57) $$H(z) = [K(z)\;\; L(z)].$$

It can be realized according to standard realization theory, and we partition the last matrix according to the partition of H. There exist therefore matrices A, B, C, D, M and N such that

(58) $$H(z) = [D \quad N] + M(z\,I - A)^{-1}\,[B\;\; C],$$

and we may chose M, A, B and C such that the system (M, A, [B C]) be minimal (i.e. completely controllable and observable). Take equality (46) which holds by hypothesis :

$$(z\,E - F)\;(D + M(z\,I - A)^{-1}\,B) = G.$$

Expand $(z\,I - A)^{-1}$ in a series in z^{-1}, and equate like powers on both sides. It comes :

power 1 : $ED = 0,$

power 0 : $EMB - FD = G,$

power - k : $(EMA - FM)\,A^{k-1}\,B = 0, \qquad k = 1, \ldots$

We do the same with equation (47). It comes

power 1 : $EN = 0,$

power 0 : $EMC - FN = 0,$

power - k : $(EMA - FM)\,A^{k-1}\,C = 0, \qquad k = 1, \ldots$

The "power 1" relations yield (52), "power 0" (51) and the second block of (50). The two "power-k" together can be written

$$(EMA - FM)\ [[B\ C]\ A[B\quad C]\ \dots\ A^{n-1}[B\quad C]] = 0.$$

Since $(A, [B\quad C])$ is taken completely controllable, the right matrix in this equality is surjective, and therefore we get the first part of (50). Straightforward calculation shows that the solutions (48) (49), subject to (50) (51) (52), satisfy (*), and similarily for the continuous case.

The strictly causal case is a specialization of this one with $D = 0$.

■

Notice also that the theorem yields

$$(z\,E - F)\,Y(z) = G\,U(z), \tag{59}$$

which is the direct z transform of (*), or Laplace transform of (**).

The rational matrix $H(z)$ of (57) can be considered as the generalized transfer function of the implicit (or generalized) system.

5.4. - Canonical form

A change of coordinates on y amounts to a right multiplication by an invertible $m \times m$ matrix Q of both E and F. (In case one is interested in an output $H\,y(t)$, H should be multiplied to the right by Q also). The system is not changed either if we replace some or all of the r equations (*) or (**) by independant linear combinations of them, i.e., if we multiply to the left E, F and G by an invertible $r \times r$ matrix P.

Therefore, two implicit systems (H, E, F, G) and (H_1, E_1, F_1, G_1), where H is an output matrix, will be said to be strictly equivalent if there exist two invertible matrices P and Q of appropriate dimension such that

$$H_1 = H\,Q,$$

$$E_1 = P\,E\,Q, \qquad F_1 = P\,F\,Q \tag{60}$$

$$G_1 = P\,G.$$

Relations (60) are precisely the definition of equivalence of the pencils $(z\,E - F)$ and $(z\,E_1 - F_1)$. We know, therefore, that by a proper choice of matrices P and Q, $(z\,E - F)$ can be brought into the canonical form described, e.g. in [9].

Let $\alpha_1(z)$ be a polynomial vector of minimum degree, say ε_1, such that

$$(z\,E - F)\ \alpha_1(z) = 0.$$

Let then $\alpha_2(z)$ be a polynomial independant of $\alpha_1(z)$, satisfying the same equality, and so on. The numbers $\varepsilon_1, \ldots, \varepsilon_q$ are the column minimal indices. Performing similarily for E' and F', we get the line minimal indices, say $\eta_1, \ldots, \eta_\ell$. The canonical form of $(z\,E - F)$ is block diagonal, made of four types of blocks.

i) blocks L_{ε_i} - To each ε_i, corresponds a block $\varepsilon_i \times \varepsilon_i + 1$ of the form

$$L_{\varepsilon_i} = \begin{pmatrix} z & -1 & 0 & \ldots & 0 & 0 \\ 0 & z & -1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & z & -1 \end{pmatrix}$$

In this basis we obviously have

$$\alpha_i(z) = \begin{pmatrix} 1 \\ z \\ \vdots \\ z^{\varepsilon_i} \end{pmatrix}$$

We make correspond to it the column $\ell_i(z)$ of $L(z)$:

$$\ell_i(z) = \begin{pmatrix} z^{-\varepsilon_i} \\ z^{-\varepsilon_i+1} \\ \vdots \\ 1 \end{pmatrix}$$

This makes up the matrix $L(z)$ of equation (47).

Writing equations (*) with this special form for E and F, we immediately see that each such block involves ε_i+1 coordinates of y. They always have a solution whatever the coefficients of G in the same lines, and the last coordinate of this subvector of y is free. It corresponds to a coordinate in N, the ε_i first corresponding to coordinates in V_*.

As a matter of fact, L_{ε_i} has the rational strictly proper right inverse

$$L^r_{\varepsilon_i} = \begin{pmatrix} z^{-1} & z^{-2} & \dots & z^{-\varepsilon_i} \\ 0 & z^{-1} & \dots & z^{-\varepsilon_i+1} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & z^{-1} \\ 0 & 0 & \dots & 0 \end{pmatrix}$$

so that whatever the corresponding lines of G, (46) will have a strictly proper solution for this block, obtained by multiplying these lines of G par $L^r_{\varepsilon_i}$ to the left.

ii) blocks L_{η_j}. To the row indices η_j, correspond blocks L_{η_j} of type $\eta_j + 1 \times \eta_j$, having the form of the transpose of a block L_ε.

Writing equations (*), with this block, we see that it involves η_j coordinates of y, but that the last line amounts to a recurrence relation between the elements of the sequence u(.). It can be satisfied for all sequences only if the corresponding lines of G are all zero, but then all these coordinates must be, and remain, zero. They correspond to coordinates in a complement of V^* in $\mathbb{R}^m$, and the requirement on G to (part of) condition (2).

Correspondingly, it is a simple task to see, thanks to the triangular form of L_{η_j}, that (46) can be satisfied, with a strictly proper block in K(z) if and only if the corresponding lines of G are null, the solution being then zero.

iii) blocks L_{μ_k}. These are square blocks, of type $\mu_k \times \mu_k$, corresponding to the infinite invariant factors of the pencil (z E - F). They are of the form

$$L_{\mu_k} = \begin{pmatrix} -1 & z & 0 & \cdots \\ 0 & -1 & z & \\ \vdots & & \ddots & \ddots \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

Again, writing the equations (*) for this block, we see that they involve μ_k coordinates of y, but depending in an anticausal way on the sequence u(.). Therefore these coordinates also correspond to a complement of V^* in $\mathbb{R}^m$, and the corresponding rows of G must be zero for a strictly causal solution to exist.

However, the dependance of y on u(.) is anticausal but not strictly. Therefore, a causal, but not strictly causal, solution may exist where the first coordinate of the corresponding subvector of y, is non zero, but all others zero. The same row in G may be non zero. This corresponds to the fact that E has a column of zeros in the first column of L_{μ_k}, and the corresponding coordinate of y is therefore in Ker E, but not in N. We recover conditions (28) and (27).

A complete information is given again looking at equation (46). As a matter of fact L_{μ_k} is invertible :

$$L_{\mu_k}^{-1} = \begin{pmatrix} -1 & -z & \cdots & -z^{\mu_k-1} \\ 0 & -1 & \cdots & -z^{\mu_k-2} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & & -1 \end{pmatrix}$$

so that the only solution of (46) for this block is $L_{\mu_k}^{-1} G_\mu$, which is anticipatory of μ_k-1 steps, unless G_μ has some zero rows. If its first row is the only non zero one, then the corresponding block of K(z) is proper, but not strictly.

iv) blocks L_{λ_ℓ}. These are square blocks, that together constitutes a characteristic matrix

$$z\,I - A_\lambda$$

where A_λ is in Jordan form, for example. This clearly corresponds to coordinates of y for which there is a unique strictly causal solution. They are therefore in V^*, but in a complement of V_*. The corresponding block of K(z) is $(z\,I - A_\lambda)^{-1}$. The corresponding eigenvalues are the essential eigenvalues of the pair (E, F).

6. - <u>CONCLUSION</u>

We have a simple theory of singular implicit systems, whether they are square, or over-or underdetermined. It should be noted that overdetermination may go along with non unicity of the solution, in a non trivial way.

The recurrence defining the various subspaces $\mathcal{V}^{*}$, $\mathcal{W}^{*}$, $\mathcal{W}_{*}$, $\mathcal{V}_{*}$, provide the basis for finite algorithms, unfortunately rather ill behaved in terms of robustness in their naive form. They involve finding zero determinants and computing right or left inverses, numerically difficult operations. Standard techniques could be applied to improve them (like computing the rank of AA^{*}, or $A^{*}A$, instead of A).

The stage seems to be set to extend a significant part of Rosenbrock's theory to these systems, and of its modern developments, in the spirit of Wolovich or Fuhrman.

A domain of interest is naturally the use of tools of two player-control systems theory to study the property of implicit systems : making an output sequence unique (decoupling v(.) through feedback), ensuring that all trajectories meet a given subspace at a given instant (capturing the state), or that some do (controllability through v), insuring that all trajectories will do better than a given amount with respect to some criterion (dynamical games), etc.

[1] G. VERGHESE, B. LEVY, T. KAILATH - Generalized State Space Systems, 1978 IEEE Decision & Control Conference, San Diego, Ca, 1979.

[2] HH. ROSENBROCK - State-space and Multivariable Theory, Nelson, London, 1970.

[3] D.G. LUENBERGER - "Dynamic Systems in Description Form", IEEE Trans. on Automatic Control, AC 22, pp 312-321, 1977.

[4] D.G. LUENBERGER - "Time Invariant Descriptor Systems" Automatica 14, pp 473-480, 1978.

[5] D.N. STENGEL, R.E. LARSON, D.G. LUENBERGER, T.B. CLINE - "A Descriptor Variable Approach to Modeling and Optimization of Large Scale Systems" Proceedings of the Engineering Fondation Conference on System Egineering for Power : organization al forms for large Scale Systems, Vol.7. DAVOS, Switzerland, 1979.

[6] P. VAN DOOREN, G. VERGHESE, T. KAILATH - Properties of the system matrix of a generalized state space system. IEEE 1978 D.C.C., San Diego, Ca. 1979.

[7] S.L. CAMPBELL - "Applications of the Drazin Inverse to linear systems of differential equations with singular constant coefficients". SIAM J[al] on Applied Mathematics, vol 31, pp 411-425, 1976.

[8] S.L. CAMPBELL - "Limit Behaviour of Solutions of Singular Difference Equations" Linear Algebra and its Applications, vol 23, pp 167-179, 1979.

[9] F.R. GANTMACHER - Théorie des matrices, tome 2. Dunod, Paris 1966 (Translated from the russian).

[10] P. BERNHARD - "Sur la commandabilité des systèmes dynamiques discrets à deux joueurs" RAIRO J3, pp 53-68, 1972.

[11] R.E. KALMAN - "Kronecker Invariants and Feedback" in Ordinary Differential Equations, 1971 NRL-MRC conference, L. Weiss ed. Academic, Paris 1972.

[12] G.D. FORNEY, Jr, - "Minimal Bases of Rational Vector Spaces, with Applications to Multivariable Linear Systems", SIAM Journal on Control, 13, pp 493-520, 1975.

[13] J.J.M. EVERS Linear Programming over an Infinite Horizon, Tilburg University Press, Academic Book Service of Holland, 1973.

[14] J.H. WILKINSON "Linear Differential Equations and Kronecker's Canonical Form", in Recent Advances in Numerical Analysis, De Boor and Golub ed., Academic Press, 1979 (Proceedings of a conference held in Madison, 1978).